

L'ogre et la toile

Le rendez-vous de l'histoire et des archives du web

Valérie SCHAFER et Benjamin THIERRY

« Le bon historien, lui, ressemble à l'ogre de la légende. Là où il flaire la chair humaine, il sait que là est son gibier », écrivait Marc Bloch dans *Apologie pour l'Histoire* (Bloch, 1997 : 4). Si les historiens affectionnent la chair humaine, ils la préfèrent généralement déjà froide. Néanmoins, ils ne pourront rester longtemps insensibles au potentiel des sources nativement numériques (en anglais *born digital heritage*), et en particulier des archives du web, en dépit de leur jeunesse. Le fait qu'elles n'aient que vingt ans d'existence explique sans doute qu'elles ne soient pas encore – ou si peu – entrées dans les pratiques¹.

1. Certains y voient même un désintérêt de l'historien pour tout ce qui relève du numérique, notamment en ce qui concerne la réflexion épistémologique : « Reste que la position sans doute la plus fréquente, et qui n'est peut-être pas la moins dommageable, est celle du désintérêt, à tout le moins celle qui consiste pour les historiens à adapter leurs pratiques aux nouveaux moyens de communication, sans s'interroger sur ce que ces derniers font à l'histoire, conçue à la fois comme savoir et profession » (Delalande et Vincent, 2011 : 5), ou encore : « De telles innovations cependant sont encore rares chez les historiens, moins enclins que les sociologues ou les géographes à explorer les possibilités de l'ordre informatique » (Rygiel, 2006 : 77).

En effet, rares sont les historien(ne)s à avoir commencé à s'en emparer et à penser leur exploitation².

Les réflexions sur les rapports entre numérique et histoire portent depuis plusieurs années sur ce que fait l'Internet à l'histoire, sur le patrimoine numérisé ou sur les outils³. Il ne s'agit pas de nier l'importance de ces réflexions et on peut même s'étonner de la rareté des débats et des prises de position de la communauté historique française sur le développement d'une *digital history*⁴, qui dépasse de loin la querelle des Anciens et des Modernes et interroge les fondements même de la discipline.

Toutefois, notre propos ici est de porter le regard vers un autre enjeu fondamental que porte en lui le numérique : la conservation de documents pléthoriques liés à l'archivage du web, réalisé à partir de 1996 par la fondation Internet Archive et depuis 2006 en France sous l'égide de la Bibliothèque nationale de France (BnF) et de l'Institut national de l'audiovisuel (INA) dans le cadre du dépôt légal⁵. Devant la masse d'informations disponibles et les nouvelles médiations qui permettent d'y accéder, il est impératif de commencer à mesurer les enjeux épistémologiques et méthodologiques sous-tendus par la promesse d'un accès renouvelé au passé grâce à ces

2. Voir Brügger (2010).

3. Citons sans volonté d'exhaustivité Clavert et Noiret (2013), Poirrier (2010), Rygiel (2005), Weller (2013).

4. « *Digital history is a rapidly changing field. New methods and formats are currently being developed. This means that "digital history" is a difficult term to define. However, it is possible to identify general characteristics. Digital history represents a democratization of history in that anyone with access to the Internet can have their voice heard, including marginalized groups which were often excluded in the 'grand narratives' of nation and empire. In contrast to earlier media formats, digital history texts tend to be non-linear and interactive, encouraging user participation and engagement* » (Arthur, 2012 : 180). « L'histoire numérique (*digital history*) est un champ en mutation rapide. De nouvelles méthodes et formats sont actuellement en cours de développement. Cela implique que le terme *digital history* est difficile à définir. Cependant, on peut en identifier des caractéristiques générales. L'histoire numérique représente une démocratisation de l'histoire en ce sens que n'importe qui ayant accès à Internet peut être entendu, les groupes marginalisés souvent exclus du grand récit de la nation et de l'empire compris. Par comparaison avec les formats médiatiques précédents, les textes de l'histoire numérique tendent à être non linéaires et interactifs, encourageant la participation des utilisateurs et leur engagement. » Notre traduction.

5. À la fin de l'année 2013, par exemple, ces archives comptaient 21,2 milliards de fichiers à la BnF.

ressources inédites. Pris entre déluge informationnel, fragmentation des sources, impossibilité de l'exhaustivité, traces partielles et gigantisme du fonds, comment l'historien peut-il envisager l'exploitation de ces archives ? Leur nature et leur nombre imposent-ils de nouvelles manières d'être au passé ? Des méthodologies sont-elles à inventer ? La discipline est-elle à l'aube d'un renouvellement profond ? C'est en tissant un dialogue avec certains textes fondateurs de l'épistémologie historique que nous esquisserons quelques pistes de réflexion concernant l'impact des archives du web sur la manière de produire et d'écrire l'histoire⁶.

Le goût de l'archive⁷

Dans les bibliothèques, le personnel (conservateurs et magasiniers) ne se perd pas en mer ; il parle d'elle [de l'archive] en nombre de kilomètres de travées qu'elle occupe. C'est une autre forme de gigantisme ou bien une astucieuse façon de l'appriivoiser tout en marquant d'emblée l'utopie que représenterait la volonté d'en prendre un jour exhaustivement possession. La métaphore du système métrique crée le paradoxe : allongée sur les rayons, mesurée en mètres de ruban comme nos routes, elle apparaît infinie, peut-être même indéchiffrable. Lit-on une autoroute, fût-elle de papier ? (Farge, 1989 : 12).

À la question fondamentale posée par Arlette Farge en 1989, le déluge informationnel promis par les archives du web ajoute une interrogation qui, si elle ne varie pas sur le fond, prend en compte le nouveau contexte s'offrant à l'historien du contemporain le plus récent : « Lit-on mieux une autoroute, fût-elle de l'information ? » La réponse apportée par les archives du web adopte d'abord le superlatif en ce qui concerne les volumes offerts : les kilomètres de linéaires sont remplacés par des milliards de pages web archivées (452 milliards par Internet Archive). Au-delà du gigantisme,

6. Il paraît utile, à l'issue de cette introduction, de souligner que notre réflexion porte sur les sources nativement numériques (qui existent uniquement sous leur forme numérique initiale) et non sur les sources et archives numérisées qui engagent d'autres problématiques.

7. Pour reprendre le titre de l'ouvrage d'Arlette Farge (1989).

ces fonds modifient-ils notre rapport à l'archive ? Jamais plus, semble-t-il, l'historien ne pourra à l'instar de Patrick Boucheron dans son *Léonard et Machiavel* écrire :

Souveraine, la petite chose trône sur un lutrin feutré, entourée de mille égards [...] frôler ce papier si souple [...] tant de fois déchiffré, retranscrit, étudié, reproduit, numérisé, diffusé, disséminé comme une nuée digitale. Il est là, présence têtue et insistante, tandis que son image numérique, infiniment démultipliée court le monde (Boucheron, 2008).

Une double rupture par le nombre et la nature de l'archive s'instaure.

(Im)Pulsion d'archive

La fluidité du texte, la fugacité des contenus et le perpétuel renouvellement du web semblent devoir vouer à l'échec toute tentative pour le saisir, comme le relève l'historien danois Niels Brügger :

Pendant les Jeux olympiques de Sydney de 2000, je voulais sauvegarder le site web du quotidien danois JyllandsPosten. Je commençai au premier niveau, avec la page d'accueil, sur laquelle on pouvait lire que la joueuse de badminton danoise, Camilla Martin, allait jouer en finale trente minutes plus tard. Mon ordinateur mit environ une heure pour sauvegarder ce premier niveau, laps de temps au bout duquel je décidai de télécharger le second niveau, les « Jeux olympiques de 2000 ». Mais sur la page d'accueil de cette section, je pouvais déjà lire les résultats de la finale de badminton (la joueuse danoise avait perdu). L'état du site web – dans son ensemble – n'était plus le même que quand j'avais commencé (Brügger, 2012 : 163).

Un des premiers effets de l'utilisation du web comme source est la confusion des rôles entre archiviste et historien. En 2000, la Library of Congress commençait tout juste un projet pilote de collecte des archives du web, un an après la Nouvelle-Zélande⁸. Aussi, l'irruption du web dans les outils du quotidien et le développement de l'accès à l'information qu'il permet

8. <<http://timeline.webarchivists.org>>.

se sont conjugués avec la conscience aiguë de son caractère éphémère et le sentiment de perte qui en découle pour attirer l'historien sur le terrain de la constitution de l'archive.

À cette volonté de garder la mémoire du réseau des réseaux s'ajoute la tentation de l'exhaustivité. En 2011, Brewster Kahle, fondateur d'Internet Archive, confie ses ambitions en matière de sauvegarde :

L'Internet Archive est une bibliothèque numérique à but non lucratif. Elle est située aux États-Unis et sa visée, à la fois sociale et technologique, est de permettre un accès universel à l'ensemble de la connaissance : tous les livres, toute la musique, toutes les vidéos, accessibles partout, par tous. Notre but est de collecter le travail de l'humanité et de le rendre accessible à ceux qui voudraient l'utiliser pour s'instruire. Notre base, c'est ce qui a été publié, c'est-à-dire les choses qui ont été pensées pour être publiques : un livre, une page web ou un billet de blog ; même les tweets... (Kahle, 2011).

Mais l'entreprise est dès ses débuts confrontée aux limites mêmes de ses visées : celle de l'impossible collecte totale et universelle d'un océan de données. Les archives composées alors sont des sources inestimables, mais aussi lacunaires, qui rendent caduque toute volonté d'établir un tableau exhaustif du développement de la toile. L'ère du choix a sonné⁹.

Qu'il crée son propre archivage ou qu'il consulte des archives collectées par d'autres, l'historien doit accepter des contraintes particulières liées à la sauvegarde du contenu mis en ligne. Les contenus sauvegardés ne sont pas une reproduction fidèle, un cliché à un instant précis d'un site dans son ensemble, en raison, d'une part des sauvegardes plus ou moins profondes qui y sont effectuées (parfois seule une collecte « de surface » est réalisée, sur les pages d'accueil ou les moins profondes du site, que l'on peut se représenter comme une succession de strates), et d'autre part du caractère rapidement changeant du web (le temps qu'un site soit collecté, ses contenus

9. Prenons l'exemple de tf1.fr : entre 1996 et 2000, 18 collectes ont été effectuées par l'Internet Archive, ainsi pour l'année 1997 les 12 février, 14 juin et 17 octobre.

ont pu évoluer¹⁰). La sauvegarde du web est par essence lacunaire. Mais cela ne diffère en rien de la sauvegarde de tout autre patrimoine humain. Quel chercheur n'a jamais pesté contre les vides d'une série d'archives, les omissions d'un document ou les discontinuités d'un texte ?

L'archive – comme l'ensemble des textes qui se trouvent en ligne – est par ailleurs rendue « fluide », insaisissable, intraversable, car s'étant multipliée, du fait de la capacité de reproductibilité à coût marginal du texte en ligne. Une infinité de versions du même texte se trouve alors accessible simultanément, variations comprises, dans une sorte de superposition des possibles que ne renierait probablement pas Hugh Everett.

Il faut ainsi composer avec la lacune comme avec la redondance, la présence de doublons ou encore de sites statiques, non mis à jour. Alors que les archivistes se demandent ce qu'il convient de faire des redondances ou doublons, on peut avancer qu'ils sont pour l'historien peut-être aussi intéressants que le perpétuel mouvement. Et s'ils ne peuvent permettre avec exactitude de dater la mort d'un site, du moins les sites figés entre plusieurs moissons donnent quelques indications sur leur destin, alors qu'il est parfois difficile pour les sites web conservés par Internet Archive de savoir quand ceux-ci ont été ouverts et si l'on a affaire à un site qui a déjà plusieurs mois ou années d'existence en amont de sa première collecte.

Serge Noiret notait déjà en 2011 que « le *digital turn* a rendu précaire un certain nombre de concepts chers aux historiens comme celui de la pérennité des sources et la capacité de reproduire dans le temps une analyse qui s'y réfère » (Noiret, 2011 : 3-4). La question de la pérennité de la source et de sa citation demeure enfin un problème pour le chercheur comme pour l'archiviste : des institutions réfléchissent à la création d'URL facilitant la citation (permalien) et qui répondent à la frustration que bien des chercheurs et leurs lecteurs ont pu rencontrer quand une référence faite à une source web ne donne plus à voir que le désarmant « 404 not found », quelques années ou même quelques mois plus tard lorsque le

10. Il n'est pas non plus rare qu'au cours de sa navigation dans une archive web le chercheur constate des sauts temporels. Ainsi la page du *Monde* dans la Wayback Machine du 21 février 1999 renvoie pour le lien « Nouvelles technologies » à celle du 8 février 1999.

lien a disparu. Si la garantie de pouvoir citer et retrouver une page web archivée constitue un réel progrès, celle de retrouver une page précise dans les archives du web reste aléatoire pour de nombreux contenus. Comme le relevait Niels Brügger :

Si un historien du web cite et veut se référer à une page web spécifique sur tf1.fr, il ne peut en aucune façon être garanti que cette page web précise sera retrouvée, soit parce qu'elle n'existe dans aucune archive, soit parce qu'elle peut être trouvée dans une archive d'apparence différente, le processus d'archivage l'ayant transformée (Brügger, 2012 : 165).

L'historien doit aussi accepter l'absence de certains éléments – des images et formats perdus – qui donnent à voir une archive parcellaire, ou des contenus sauvegardés qui ne sont pas ceux qui feraient sens à ses yeux. Là encore si les réflexions sur les formats ou l'émulation sont bien engagées par les institutions de conservation, les collectes générales effectuées dans le web au titre de l'archivage fonctionnent non selon l'intérêt (toujours hypothétique et subjectif) des contenus, mais par moissonnage à intervalles temporels réguliers par des robots.

Cet état d'incertitude généralisé que certains déplorent n'est toutefois pas entièrement nouveau pour l'historien. En effet, l'ère de l'imprimé, après celle du manuscrit, a déjà fait connaître duplicatas, emprunts et autres mouvements intertextuels qui rendent le travail de repérage, d'établissement de la véracité et de la généalogie d'un texte consubstantiel à la pratique historique. Internet ne bouleverse pas radicalement la donne en la matière, mais accélère une tendance. Limite évidente pour l'histoire, les lacunes sont néanmoins inhérentes à l'idée même d'archive qui implique un choix sur ce qu'il s'agit de sauver et ce qu'il importe de laisser disparaître.

Les périphéries de l'archive

Si les images, les témoignages oraux, les sources imprimées, manuscrites, officielles ou non, ont depuis longtemps obligé l'historien à démystifier la source, le prestige du chiffre, de la parole, du sceau ou de la pellicule, les archives du web semblent peut-être encore davantage échapper à la critique d'authenticité, à la certification, aux méthodes traditionnelles de critique interne et externe du document.

Mais le caractère particulier de l'archive du web, notamment lié à sa nature numérique et au code, permet d'envisager de nouvelles manières d'accéder aux contextes de production et de diffusion de l'information. Les métadonnées ouvrent en effet des perspectives réjouissantes pour l'historien et viennent renouveler les possibilités de critiques interne et externe patiemment élaborées comme l'un des fondements de la discipline. La date, le lieu, l'auteur d'un cliché sont aujourd'hui stockés couramment dans les fichiers photographiques numériques¹¹. Il en va de même pour les textes ou les courriels. Les politiques d'archivage réservent de plus en plus une place à ces « périphéries de l'archive » : au sein de l'International Internet Preservation Consortium (IIPC)¹² ou de l'Internet Memory Research¹³ on s'interroge sur les moyens de pérenniser l'archive et de la consulter, mais également sur les possibilités d'utilisation et de rationalisation des corpus grâce aux métadonnées.

Si l'entrelacement de la critique interne et de la critique externe porte sur des critères nouveaux (point n'est besoin de développer ici la caducité de la critique externe sur le papier, l'encre, l'écriture que proposaient Charles-Victor Langlois et Charles Seignobos – Seignobos et Langlois, 1898), la méthode reste toujours valable. La date de rédaction d'un texte, de la prise d'une photographie, éventuellement la localisation de sa production ou encore les formats utilisés forment les premiers éléments d'une *littérature numérique* historique qui commence à se construire, mais qu'il reste à formaliser. Celle-ci peut s'appuyer sur les réflexions d'autres disciplines à l'instar de ce qui est mené dans le domaine de la diplomatique. Nous renvoyons en particulier à la réflexion précoce engagée par Luciana Duranti et au projet InterPARES, groupe de recherche internationale sur l'authenticité et la pérennité des documents dans les systèmes électroniques, fondé en 1999, ainsi qu'à l'article d'Anne-Marie Chabin « Peut-on parler de diplomatique numérique ? » (Chabin, 2013).

11. La spécification de format de fichier Exif (*exchangeable image file format*) est un bon exemple de cette production de données exploitables par l'historien en matière d'iconographie (voir : <<http://owl.phy.queensu.ca/~phil/exiftool/TagNames/EXIF.html>>).

12. <<http://netpreserve.org>>.

13. Voir par exemple la plateforme [m]ignify, <<http://mignify.com>> et le site de l'Internet Memory Foundation : <<http://internetmemory.org/en/>>.

À travers la multitude d'indices disséminés dans les sites, l'historien peut toujours espérer trouver des éléments de repérage (dates de mise à jour, compteurs de fréquentation, référencement dans les annuaires du web type Yahoo, etc.). Il peut toutefois avoir des difficultés à retracer la nature et les caractéristiques d'un site en partant du site même (nous verrons qu'il peut disposer d'autres ressources, *via* des archives institutionnelles, des entretiens oraux, etc.), la réalisation collective, parfois externalisée, ne permettant pas non plus d'en identifier aisément les auteurs.

Quant aux destinataires et récepteurs, ils sont encore plus difficilement perceptibles. En effet, aujourd'hui encore il semble que les études sur les publics et leur connaissance approfondie sont encore tâtonnantes dans beaucoup de services, par exemple de médiation institutionnelle, scientifique ou culturelle. Toutefois à la suite d'Antoine Prost, qui notait que « ni la critique des témoignages oraux ni celle des photographies ou des films ne diffèrent de la critique historique classique. C'est la même méthode appliquée à d'autres documents. Elle utilise parfois des savoirs spécifiques [...] » (Prost, 2010 : 66-67), on peut ajouter que la critique des sites web, si elle se doit d'employer des savoirs spécifiques, ne diffère en rien par ses ambitions de celle appliquée à d'autres documents. Elle reste au cœur de la méthode de l'histoire.

« Témoignages volontaires et involontaires¹⁴ »

La critique externe se doit d'être adaptée aux archives du web ; de même, la critique interne se heurte à de nouvelles exigences impliquant de dépasser certaines catégories qui ont pu faire preuve de leur efficacité par le passé. Ainsi en va-t-il de la critique d'exactitude ou de sincérité. En effet, on peut être tenté de l'appliquer à un site par exemple de presse en ligne – et encore seulement partiellement dans la mesure où les commentaires et annotations des lecteurs peuvent difficilement entrer dans la même catégorie d'analyse que l'article d'un journaliste –, mais la distinction classique entre témoignages volontaires et involontaires semble plus délicate. Faut-il reprendre et affiner cette typologie ? Peut-elle maintenir la césure entre des sources narratives destinées à nous informer et « ces indices que, sans préméditation,

14. Prost (2010 : 62).

le passé laisse tomber sur la route » (Bloch, 1997 : 25) ? Faut-il considérer toute trace sur le web comme volontaire, distinguer l'agentivité humaine et celle des machines, repenser à l'aune de l'extimité la démarcation entre privé et public ? Il convient avant tout de se demander si distinguer des expressions officielles ou non, des sources marquées par l'anonymat ou le pseudonymat de celles dont l'auteur est identifiable fait sens à l'heure des identités numériques multiples. Est-ce encore adapté aux multiples identités numériques qu'endosse une personne au fil de ses navigations et formes d'expression sur le web ? Le support lui-même ne peut jouer le rôle de discriminant : ainsi, un gazouillis posté sur Twitter par Valérie Trierweiller concernant Ségolène Royal et celui d'un anonyme n'ont ni la même fonction, ni les mêmes répercussions. Ici encore, l'appréciation des contextes vient confirmer l'histoire dans son rôle de science de l'unique, du présent conjugué au passé, qui ne se satisfait qu'imparfaitement de certains modèles et catégorisations ambitieuses.

Enfin, on peut se demander s'il en ira réellement autrement des archives du web que de celles qui les ont précédées : tout fonds d'archives donne à voir par la sélection des documents qui s'y trouvent et les formes du classement qui y sont adoptées une expression du pouvoir qu'exercent les individus et les groupes sur les traces laissées à la postérité. « L'histoire est toujours faite par les vainqueurs », dit l'adage populaire. Selon des géométries spécifiques aux sujets abordés, les traces laissées en ligne et les acteurs de l'histoire mis au premier plan pourraient être le résultat de la capacité de certains à capter la lumière en laissant leurs adversaires dans l'ombre.

« L'histoire comme compréhension¹⁵ »

Rien de ce que nous avons vu jusqu'ici ne nous donne de l'histoire une idée claire. Elle semble perpétuellement occupée à concilier des contradictions. Il lui faut des faits, tirés de sources, mais sans questions les traces restent muettes et ne sont mêmes pas « sources ». [...] Quant au temps, à la dimension diachronique constitutive de la question historique, ce n'est pas un cadre vide que l'on remplirait de faits,

15. Prost (2010 : 145 sq.).

mais une structure façonnée par la société et par l'histoire déjà écrite. L'historien qui le travaille comme un matériau doit le considérer aussi comme un acteur à part entière de son scénario (Prost, 2010 : 145 sq.).

« À chaque phénomène sa périodisation¹⁶ »

Comme l'ont relevé notamment Marc Bloch et Antoine Prost, chaque phénomène historique a sa périodisation propre et il convient pour l'historien de démêler l'écheveau chronologique pour construire le temps propre de son objet. Les archives du web ne font pas exception à la règle : une simple page dédiée à l'abdication du pape Benoît XVI, selon qu'elle est convoquée par un historien de la papauté, un historien de la presse religieuse ou de la communication en ligne, est travaillée, analysée différemment, mais aussi insérée dans une périodisation propre à l'objet d'étude. « L'historien ne reconstruit pas la totalité du temps à chaque recherche : il reçoit un temps qui a déjà été travaillé, déjà périodisé par d'autres historiens » (*ibid.* : 116). Mais dans le cas des archives du web ne faut-il pas aussi tenir compte d'un temps du web, en termes de réception, de pratiques ou d'usages, qui entre en écho avec le temps de l'objet étudié ? Quels sont les événements signifiants d'une histoire de la toile et comment croiser son temps avec celui de l'étude historique ?

Les contextes, notamment techniques, de développement du web semblent en effet profondément influencer l'information dans sa globalité. Des balbutiements dans les années 1990, avec une acculturation et une appropriation renforcée du milieu des années 1990 au milieu des années 2000, jusqu'à la rupture dite « 2.0 », les possibilités de production, d'édition et de réception de l'information ont fortement et rapidement évolué. La consommation facturée selon le temps de connexion puis libérée avec l'apparition des forfaits haut débit à la fin des années 1990, l'accélération constante des vitesses de consultation du contenu, l'enrichissement de ce dernier, notamment en images, vidéos et animations, bouleversent les contextes informationnels que l'historien doit prendre en compte pour s'assurer d'une réelle compréhension de la place de l'archive. Les rythmes de l'usage sont également fondamentaux dans l'acquisition d'une réelle

16. *Ibid.* : 118.

littératie numérique historienne : l'irruption si souvent célébrée de l'amateur, la place des contenus générés par l'utilisateur et, plus globalement, la question de la participation du public au média forment des structures heuristiques nécessaires à l'historien.

Des histoires nationales et transversales du web, encore à écrire, seraient à cet égard utiles pour offrir un socle commun aux praticiens de l'histoire par le web ; ils pourraient y puiser des repères en matière de publics, de conditions de production, de lignes générales de développement et de politiques (comme en fournit, par exemple, le Plan d'action gouvernemental pour la société de l'information lancé en 1998 par le gouvernement de Lionel Jospin pour la France). En matière également économique (structure de l'offre des fournisseurs d'accès à Internet [FAI]), technique (passage à l'ADSL, offres 3G, forfaits), ou encore de contenus (lancement des premiers sites d'e-commerce en France à la fin des années 1990 – iBazar et Cdiscount en 1998, Aucland, Kelkoo et RueDuCommerce en 1999, Pixmania, Amazon.fr, PriceMinister, Voyages-sncf.com et Assurland en 2000, etc.). Comment ne pas risquer l'anachronisme en pensant le passé avec des concepts contemporains, la « révolution » en permanence aux lèvres, lorsque l'on est inattentif d'une part aux métamorphoses de l'Internet et du web en seulement quelques décennies, d'autre part aux systèmes de représentation qui leur sont propres ? Les valeurs attachées à l'interaction ou à la participation ont connu des évolutions qui dépassent le web. Les formes d'expression sur la toile sont solidaires de représentations sociales façonnant le support et renvoient à des « imaginaires de l'Internet » qui doivent s'analyser dans le temps court de l'histoire de l'Internet, dans le temps long de l'histoire des écritures (Herrenschmidt, 2007) d'une manière singulière et non figée. Ces vœux ne sont finalement qu'un appel à la longue durée, d'une part, à une perception globale des objets étudiés, de l'autre.

De l'histoire bataille à l'histoire pagaille

Au cœur de la grande critique portée par Marc Bloch et Lucien Febvre à l'encontre de leurs prédécesseurs méthodiques se trouve le reproche de l'établissement d'une « histoire bataille », sensible aux grands événements, aux têtes couronnées plus qu'aux mouvements d'ensemble, économiques et sociaux en tête, qui animent l'histoire et ses convulsions (Le Goff, 1999). Avec Fernand Braudel et sa *Méditerranée* (Braudel, 1966), la longue durée

devient l'horizon indépassable de la recherche historique et ouvre des perspectives qui sont encore vivaces aujourd'hui et que l'on peut déceler aux détours des attentes portées par l'histoire globale par exemple.

Fondamentalement, de la fin des années 1920 aux années 1980, l'histoire se décloisonne, prend de l'altitude et embrasse de nouvelles échelles de compréhension en établissant dans le même temps des dialogues féconds avec les autres disciplines des sciences humaines et sociales. Ce décloisonnement n'a pas pour autant poussé l'historien à abandonner les terrains qui étaient les siens, mais a donné à voir une absorption continue de nouveaux objets, de nouvelles échelles, de nouvelles méthodologies.

Le web en tant que nouveau territoire d'histoire est l'occasion de poursuivre ce décloisonnement. Chambre d'écho du monde en tant que média, moyen d'action en tant qu'outil de communication, de mobilisation et parfois d'affrontement, il n'est pas séparé du réel, il n'est pas virtuel, comme certains ont pu le penser dans le cadre d'une dichotomie plaisante mais tout à fait illusoire.

Aussi, pour que l'ogre se repaisse de ce qu'il trouve sur la toile sans s'y empêtrer, il faut éviter de faire du web l'alpha et l'oméga d'une nouvelle histoire borgne qui ne prendrait que la partie émergée du numérique comme objet ou qui ne verrait les événements que par le petit bout de la lorgnette du réseau des réseaux. Au même titre que la presse, l'archive diplomatique ou tout autre type de document, l'analyse par l'archive web ne peut être féconde que croisée, c'est-à-dire associée à d'autres. Le web doit s'intégrer dans une visée globale de compréhension des phénomènes et ne pas devenir hégémonique dans les pratiques, sous prétexte de nouveauté ou de facilité d'accès.

Peu de phénomènes font depuis vingt ans l'économie de laisser des traces en ligne, l'historien doit bien sûr les exploiter. Pour ne citer qu'un exemple, il est nécessaire que l'histoire des entreprises adopte rapidement l'analyse de la communication en ligne dans le spectre de ses préoccupations. Quelles formes prend-elle ? Y décèle-t-on les mutations organisationnelles récentes ? Comment le consommateur, l'usager, l'utilisateur y concourent, la subissent ou la ressentent ? Autant de problématiques que l'archive du web doit permettre d'éclairer.

Bien entendu, l'analyse de ces objets nouveaux ne peut négliger pour autant les archives plus traditionnelles. Comment contextualiser la communication en ligne sans les apports de l'archive papier (bilans comptables,

rapports d'activité, etc.), du témoignage lorsque les acteurs de l'histoire immédiate sont encore en vie, ou de la presse ?

Quand bien même l'objet serait lui-même numérique, ses traces numériques ne suffisent pas. L'histoire même d'Internet et du web ne peut se passer de sources qui ne constituent pas directement la mémoire du réseau. Ainsi, comment analyser les sites web des années 1990 sans chercher quelles sont alors les conditions de navigation, le taux d'équipement, le type de ces équipements et les attentes d'un grand public en gestation ? Il suffit de rappeler qu'en 1996 la France compte seulement 4 234 noms de domaines (contre 362 388 aux États-Unis) pour éclairer cette remarque du site Statisticator au sujet de celui dédié à la Fête de l'Internet : « Pathétique. Que dire d'autre ? Le site de la Fête de l'Internet a des stats pour le moins dramatiquement basses. Reflètent-elles l'état du Net ? En tout cas, elles inspirent encore une fois cette angoissante question : si dans le cyber-espace personne ne sait que vous êtes un chien, n'est-ce pas simplement parce que dans le cyber-espace, il n'y a personne ? » Il suffit de revenir à *La carte française des inforoutes* (Curtil, 1996), ouvrage dans lequel on peut lire : « Un utilisateur moyen devrait passer près d'une heure par jour sur Internet. Cette durée peut paraître longue mais le réseau est tellement encombré qu'il faut entre 10 et 20 minutes au minimum pour trouver et charger le document souhaité », pour mesurer la distance qui nous sépare déjà du web du milieu des années 1990.

En ce sens, le « *web historian* » n'existe pas, il n'y a que des historiens tout court qui s'affrontent à de nouvelles sources et à de nouvelles manières de les rassembler, de les analyser et d'en rendre compte ; des historiens qui pratiquent une « histoire pagaille » par sa diversité et sa richesse plutôt qu'une « histoire bataille » ripolinée aux couleurs du numérique qui ne serait pas sans rappeler les errements passés de l'histoire internaliste des techniques (Griset et Bouvier, 2012).

17. <http://www.kitetoa.com/Pages/Statisticator/Pages/Pages/les_news.htm>.

« Il n'y a pas davantage de document sans question¹⁸ »

Affirmant avec force que c'est la question posée par l'historien qui érige les traces laissées par le passé en sources et documents, Antoine Prost, comme Marc Bloch ou Henri-Irénée Marrou (1975) avant lui, rappelle le primat de la problématique sur la source qui n'est jamais un donné immédiat du réel. Dès lors il s'agit d'examiner la constitution des corpus numériques et, en particulier, ce que fera l'historien de ceux constitués par les institutions d'archivage. Ainsi la BnF réalise-t-elle, à côté des collectes courantes qui portent sur une dizaine de milliers de sites, archivés depuis 2011 à des fréquences variables allant d'« une fois par semaine » à « une fois par an », des collectes « projets », qui « se caractérisent par leur sensibilité plus forte à l'actualité et par leur transversalité. Elles sont souvent réalisées en coopération avec des partenaires externes (bibliothèques, centres de recherches, associations)¹⁹ ». En 2013, ont ainsi été sauvegardés une centaine de sites d'actualité capturés tous les jours, une sélection de vidéos de Dailymotion, 450 sites illustrant l'expression intime et littéraire en ligne – cette dernière collecte ayant été réalisée en partenariat avec l'Association pour l'autobiographie –, 570 sites sur le thème des solidarités, 230 sites donnant à voir la diversité des mouvements contemporains, les pratiques et le sens de l'activité contestataire, 800 sites de l'État, des administrations territoriales ou des organes européens et internationaux, ou encore un échantillonnage de 200 000 blogs à partir de 20 plateformes. La tentation peut être forte de privilégier des secteurs de la vie numérique davantage conservés que d'autres.

La « BnF dispose par ailleurs d'une procédure de "collecte d'urgence" qui permet de capturer rapidement des sites à collecter à une date précise [...] car susceptibles de disparaître²⁰ ».

Connaître les modalités et les choix qui président à l'archivage, s'approprier des corpus conçus par d'autres acteurs (dans les cas précédemment

18. Prost (2010 : 81).

19. <http://www.bnf.fr/fr/collections_et_services/anx_pres/a.collectes_ciblees_arch_internet.html>.

20. *Ibid.* Cette collecte d'urgence a notamment été activée en janvier 2015 à la suite des attentats commis à *Charlie Hebdo* et à la porte de Vincennes à Paris.

évoqués, le travail de patrimonialisation et de préservation effectué par la BnF en lien avec l'Association pour l'autobiographie ; dans le cadre du projet e-Diasporas²¹ mené sous la direction de Dana Diminescu, la préservation de centaines de sites diasporiques avec l'INA) impliquent de pénétrer dans la fabrique de l'archive, pour ne pas en être esclave et pour ne pas renouer avec la pratique de la source imprimée et éditée qui masque la source originelle. Le projet e-Diasporas en documentant la constitution de ses corpus offre ainsi une approche originale utile à leur réappropriation par d'autres chercheurs.

Pénétrer la fabrique de l'archivage du web est d'autant plus nécessaire que celui-ci a évolué rapidement depuis 1996 en lien non seulement avec les capacités techniques, les mesures institutionnelles et législatives à l'image du dépôt légal, mais aussi avec les modèles épistémologiques de la recherche. Ainsi Louise Merzeau, lors de l'assemblée générale de l'IIPC en mai 2014²², distingue trois phases de l'archivage du web : à une première calquée sur le modèle de la bibliothèque, où la collecte est souvent de surface, puis une deuxième fondée sur celui de l'*exemplar* où s'affirment l'enjeu mémoriel et le « carottage » au sein de certains sites (archiver tous les sites se rapportant à un sujet, un événement, un contenu), succéderait aujourd'hui le modèle du web temporel, tourné vers davantage de plasticité, constituant des entités inédites, hybrides renvoyant à des temporalités différentes. Ce passage d'un web statique à un web temporel implique de ne plus considérer l'archive comme une entité close, mais de penser un système, des environnements complexes faits de circulation, de mouvements et d'ajustements informationnels en flux et reconfiguration permanents. Soulignons avec Claire Lemercier que « l'analyse de réseau n'est pas un outil miracle pour penser le lien social, du passé ou du présent, mais [...] peut aider à conclure sur des hypothèses qui doivent être élaborées au préalable et à partir de sources classiques, simplement envisagées de façon nouvelle. Cette question de sources demeure le principal obstacle à un usage largement répandu de la technique : la première tâche de l'historien reste de comprendre "ce que

21. <<http://www.e-diasporas.fr>>.

22. <[http://netpreserve.org/sites/default/files/attachments/IIPC_GA2014 - Open Day schedule v2_0.pdf](http://netpreserve.org/sites/default/files/attachments/IIPC_GA2014_-_Open_Day_schedule_v2_0.pdf)>.

les sources interdisent de faire” (Boutier, 2003) » (Lemerrier, 2005). Il est possible d’imaginer son application aux archives de la toile pour renforcer l’appréhension de la plasticité du web et de ses réseaux, à condition toutefois que derrière ces réseaux, les graphes, cartographies et la prouesse des outils, se donnent à voir les hommes : « Derrière les traits sensibles du paysage, les outils ou les machines, derrière les écrits en apparence les plus glacés et les institutions en apparence les plus complètement détachées de ceux qui les ont établies, ce sont les hommes que l’histoire veut saisir. Qui n’y parvient pas ne sera jamais, au mieux, qu’un manœuvre de l’érudition », précisait Marc Bloch (1997 : 4). Ce point n’est pas anecdotique et constitue le soubassement d’un projet scientifique (faire de l’histoire une science de l’homme et non pas uniquement du passé) et la fondation d’un horizon éthique : l’histoire est un humanisme, car elle place l’homme au cœur de ses préoccupations.

« L’historien de demain sera programmeur ou il ne sera plus²³ » ?

Lawrence Stone, fin connaisseur de l’école des Annales, s’était montré dès la fin des années 1960 très critique à l’égard des espoirs portés par les études historiques quantitativistes, leurs débauches de moyens et la transformation de l’historien en informaticien zélé qui pense trouver des réponses dans les procédures de traitement des données plus que dans l’analyse des sources (Stone, 1979). Emmanuel Le Roy Ladurie, au terme d’une enquête sur les conscrits français au XIX^e siècle, après avoir, exprimé son enthousiasme sur les perspectives qu’offrait l’ordinateur à l’historien, partit pour Montaignou, village occitan, dont il retraça les vies intimes et toute une existence quotidienne en s’éloignant des mesures et des statistiques, comme se plaît à le rappeler Antoine Prost... Faut-il y voir une métaphore des attentes et, finalement, de la déception qu’apporte l’informatique à l’historien ? Un présage pour ce qui relève aujourd’hui de l’utilisation des archives du web ?

23. Le Roy Ladurie (1973 : 14).

Si les promesses des outils et des archives numériques laissent aux historiens grande ouverte la porte en matière de méthodes et ne présagent en rien des approches qu'ils privilégieront (qualitatives ou quantitatives pour peu que ces binarismes puissent encore être valides), l'histoire n'est pas aujourd'hui plus qu'hier à l'abri de la tentation du scientisme de la quantité, des données et du penchant pour la *data-driven science* qui a pu faire dire dans d'autres champs scientifiques que l'on assistait à la fin des théories (voir Anderson, 2008). Le déluge informationnel qui menace également les historiens pose la question de la manière dont ils sauront continuer à allier les deux éléments de toute pensée historique selon Robin G. Collingwood : données et principes d'interprétation, qui n'existent pas séparément – l'historien ne pouvant pas récolter les données dans un premier temps, pour les interpréter ensuite²⁴. Entre subjectivisme et exigence scientifique, échantillonnage ou prétention à l'universalité et à la totalité, va-t-on, sous couvert de l'utilisation de nouvelles archives, voir renaître d'anciennes querelles ou au contraire les héritages de l'historiographie permettront-ils de dépasser ces dichotomies ?

Alors que François Dosse soulignait dans « L'histoire en miettes » qu'un « effet, pervers, de l'utilisation et du fétichisme de l'ordinateur est de privilégier la répétition de phénomènes de même nature, donc la longue durée permanente et immobile, ce qui aboutit à l'élimination de l'événement, de la rupture signifiante » (Dosse, 1985 : 59), les archives du web offrent une occasion de réexaminer le rapport à l'ordinateur, au numérique, à ses outils, mais aussi à ses contenus. Faut-il réinventer le métier (Genet et Zorzi, 2011) ? Pour le moins faut-il repenser leur relation sans craindre la crise identitaire ? Antoine Prost l'avait évoqué en son temps en forme de bilan pour le projet des Annales :

La prétention scientifique que partageaient, malgré leurs désaccords, un Seignobos et un Simiand vacille sous les coups d'un subjectivisme qui annexe l'histoire à la littérature [...] l'entreprise unificatrice de Braudel et des partisans d'une histoire totale qui récapitulait en elle l'apport de toutes les autres sciences sociales a débouché sur une crise de confiance : à force d'emprunter à l'économie,

24. Robin G. Collingwood, cité in Prost (2010 : 81).

à la sociologie, à l'ethnologie, à la linguistique leurs questions, leurs concepts et leurs méthodes, l'histoire connaît aujourd'hui une crise d'identité qui suscite la réflexion » (Prost, 2010 : 10).

Aujourd'hui, si le temps des emprunts se poursuit, on ne peut que souhaiter que se renforce celui du dialogue. Non seulement avec les sœurs de l'histoire, sociologie, ethnologie, anthropologie, etc., mais également avec ses cousines éloignées, les sciences dites « dures ». Ces dernières fournissent des éclairages pratiques et épistémologiques de premier ordre sur des outils qui se multiplient et dont l'historien aurait tort de se priver. Plus que le risque de crise identitaire, plus que l'« emprunt » à d'autres disciplines, se pose la question d'une véritable interdisciplinarité autour du web et de ses archives que l'histoire ne doit pas garder sous son seul empire.

Pour ce qui est de la formation, s'il est illusoire de faire de l'historien un informaticien – à peu près autant que l'inverse –, les nouvelles littératies devraient intégrer les parcours de formation, notamment doctoraux. En effet, nous avons tenté de le montrer, les archives du web ne provoquent pas de rupture épistémologique réelle et s'intègrent dans une tradition disciplinaire qui a su acclimater avec profit les précédents médias au régime de l'archive (radio, cinéma ou télévision). Elles imposent cependant une adaptation méthodologique de la pratique historique, notamment en ce qui concerne l'exploitation de leurs « périphéries ».

Quant à savoir s'il existe « une histoire numérique, souvent 2.0, pour un vaste public et une histoire faite en usant de médias traditionnels pour le seul public universitaire » (Noiret, 2011), force est de constater que la confusion actuelle entre historiens de métier et amateurs d'histoire rejoue la vieille querelle entre histoire et mémoire, entre publics passifs et sachants. Il s'agit ici surtout de l'éternel débat autour de la verticalité d'une activité qui est questionnée par l'élargissement de la participation à des artefacts de savoir au sein desquels Wikipédia joue un grand rôle. Comme le remarquent Nicolas Delalande et Julien Vincent (2011), « le principal enjeu de l'histoire digitale est sans doute la redéfinition du rôle social de l'historien, et de sa place dans la construction des discours sur l'histoire » et non une redéfinition de la pratique ou des finalités de la discipline.

Références bibliographiques

- ANDERSON, Chris**, 2008,
« The end of theory: The data deluge makes the scientific method obsolete », *Wired Magazine*, vol. 16, n° 7. Consultable en ligne : <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>.
- ARTHUR, Paul Longley**, 2012,
« Toward a global digital history », in Rohit Chopra et Radhika Gajjala (éd.), *Global Media, Culture and Identity: Theory, Cases, and Approaches*, New York, Routledge, p. 175-187.
- BLOCH, Marc**, 1997 [1949],
Apologie pour l'histoire ou Métier d'historien, Armand Colin, Paris.
- BOUCHERON, Patrick**, 2008,
Léonard et Machiavel, Paris, Verdier.
- BOUTIER, Jean**, 2003,
« Sources, objets, outils... Quelques remarques pour éviter de conclure », in Pierre-Yves Beaurepaire et Dominique Taurisson (dir.), *Les Ego-documents à l'heure de l'électronique. Nouvelles approches des espaces et réseaux relationnels*, Montpellier, Université Montpellier III, p. 535-544.
- BRAUDEL, Fernand**, 1966 [1949],
La Méditerranée et le monde méditerranéen à l'époque de Philippe II, Armand Colin, Paris.
- BRÜGGER, Niels (éd.)**, 2010,
Web History, New York, Peter Lang.
- , 2012, « L'historiographie de sites web : quelques enjeux fondamentaux », *Le Temps des médias*, n° 18, p. 159-169.
- CHABIN, Anne-Marie**, 2013,
« Peut-on parler de diplomatie numérique ? », in Valentine Frey et Matteo Trelean (dir.), *Vers un nouvel archiviste numérique*, Paris, L'Harmattan. Consultable en ligne : <<http://www.marieannechabin.fr/diplomatique-numerique/>>.
- CLAVERT, Frédéric et NOIRET, Serge**, 2013,
L'histoire contemporaine à l'ère numérique, Bruxelles, Peter Lang.
- CURTIL, Cédric**, 1996,
La carte française des inforoutes, Paris, Hermès.
- DELANDE, Nicolas et VINCENT, Julien**, 2011,
« Portrait de l'historien-ne en cyborg », *Revue d'histoire moderne et contemporaine*, n° 58, p. 5-29.
- DOSSE, François**, 1985,
« L'histoire en miettes : des Annales militantes aux Annales triomphantes », *Espaces Temps*, vol. 29, n° 29.
- FARGE, Arlette**, 1989,
Le goût de l'archive, Paris, Éditions du Seuil.
- GENET, Jean-Philippe et ZORZI, Andrea**, 2011,
Les historiens et l'informatique : un métier à réinventer, Rome, École française de Rome.
- GRISET, Pascal et BOUVIER, Yves**, 2012,
« De l'histoire des techniques à l'histoire de l'innovation. Tendances de la recherche française en histoire contemporaine », *Histoire, économie & société*, vol. 2, p. 29-43.
- HERRENSCHMIDT, Clarisse**, 2007,
Les trois écritures : langue, nombre, code, Paris, Gallimard.

KAHLE, Brewster, 2011,

« Internet Archive, le meilleur du web est déjà perdu », entretien avec Xavier Delaporte, recueilli et traduit par Thibault Henneçon pour Place de la Toile, mars. Consultable en ligne : <<http://www.internetactu.net/2011/06/28/brewster-kahle-internet-archive-le-meilleur-du-web-est-deja-perdu/>>.

LE GOFF, Jacques, 1999,

« Les "retours" dans l'historiographie française actuelle », *Les Cahiers du Centre de recherches historiques* (en ligne), n° 22. Consultable en ligne : <<http://ccrh.revues.org/2322>>.

LEMERCIER, Claire, 2005,

« Analyse de réseaux et histoire », *Revue d'histoire moderne et contemporaine*, vol. 52, n° 2, p. 88-112. Consultable en ligne : <<http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2005-2-page-88.htm%23no70>>.

LE ROY LADURIE, Emmanuel, 1973,

« L'historien et l'ordinateur », in *Le territoire de l'historien*, t. I, initialement publié dans *Le Nouvel Observateur*, 8 mai 1968.

MARROU, Henri-Irénée, 1975 [1954],

De la connaissance historique, Paris, Éditions du Seuil.

NOIRET, Serge, 2011,

« Y a-t-il une histoire numérique 2.0 ? », in Jean-Philippe Genet et Andrea Zorzi, *Les historiens et l'informatique : un métier à réinventer*, Paris, p. 235-288.

POIRRIER, Philippe, 2010,

« Internet et les historiens », in Christian Delacroix, François Dosse, Patrick Garcia, Nicolas Offenstadt (éd.), *Historiographies*, vol. 1 : *Concepts et débats*, Paris, Gallimard, p. 468-475.

PROST, Antoine, 2010 [1996],

Douze leçons sur l'histoire, Paris, Éditions du Seuil.

RYGIEL, Philippe, 2005,

« Les historiens dans l'espace électronique », in Philippe Rygiel et Serge Noiret (dir.), *Les historiens, leurs revues et Internet*, Publibook Universités, p. 9-17.

—, 2006, « L'ordinateur, le réseau et l'écriture de l'histoire », *Matériaux pour l'histoire de notre temps*, n° 82/2, p. 75-79.

SEIGNOBOS, Charles

et LANGLOIS, Charles-Victor, 1898, *Introduction aux études historiques*, Félix Alcan, Paris.

STONE, Lawrence, 1979,

« The revival of narrative: reflections on a new old history », *Past and Present*, n° 85, novembre, p. 3-24.

WELLER, Toni, 2013,

History in the Digital Age, Londres, Routledge.